ELSEVIER

Int. J. Production Economics 80 (2002) 201–212

international journal of production economics

www.elsevier.com/locate/dsw

Transshipments: An emerging inventory recourse to achieve supply chain leagility

Yale T. Herer[a,*,1], Michal Tzur[b], Enver Yücesan[c]

[a] Faculty of Industrial Engineering and Management, Technion-Israel Institute of Technology, 32000 Haifa, Israel
[b] Department of Industrial Engineering, Tel Aviv University, Tel Aviv, Israel
[c] Technology Management Area, INSEAD, Fontainebleau, France

Received 6 May 2001; accepted 3 May 2002

Abstract

Supply chain designs are constrained by the cost-service trade-off. Cost minimization typically leads to physically efficient or lean supply chains at the expense of customer responsiveness or agility. Recently, the concept of leagility has been introduced. Research on leagility, defined as the capability of concurrently deploying the lean and agile paradigms, hinges heavily on the identification of the decoupling point, which, in turn, is enabled by postponement. Postponement strategies, however, present a cross-functional challenge for implementation. As a tactical solution to achieve leagility without postponement, we introduce transshipments, which represent a common practice in multi-location inventory systems involving monitored movement of stock between locations at the same echelon level of the supply chain. Through a series of models, we establish how transshipments can be used to enhance both agility and leanness. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Transshipments; Supply chain management; Leagility; Inventory control; Newsvendor model

1. Motivation

A supply chain is a network consisting of suppliers, manufacturers, distributors, retailers, and customers. Supply chains perform two principal functions (Fisher, 1997): the *physical* function of transformation, storage and transportation, and the *market mediation* function of matching demand and supply in an uncertain and dynamic environment. While the physical function has been extensively studied within the production control and inventory management literature, innovative approaches have recently been emerging to the market mediation function. These approaches are classified in Fig. 1. This paper goes beyond these approaches and introduces the strategy of transshipments to support both the physical and market mediation functions of the supply chain.

Supply chain design is typically thought of as a network configuration, that is, the specification of customer zones, selection of manufacturing and distribution facilities, and allocation of product families to these sites. There exists a wealth of operations research tools for generating optimal

---

*Corresponding author. Tel.: +972-4-829-4423; fax: +972-4-823-5194.

E-mail address: yale@technion.ac.il (Y.T. Herer).

[1] This research was performed, in part, while the author was at the Department of Industrial Engineering, Tel Aviv University.

PII: S0925-5273(02)00254-2

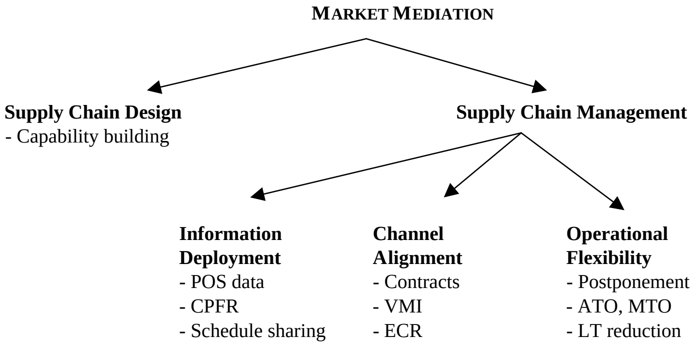

Fig. 1. Matching demand and supply in a supply chain.

network topologies (for example, see Camm et al., 1997). According to Fine (2000), however, supply chain design is also concerned with the prioritization of the capabilities to be developed and retained internally and the forging of new partnerships with other entities along a supply network. In particular, supply chain design ought to be thought of as a *dynamic* process of assembling chains of capabilities and not just collaborating organizations.

Supply chain management, on the other hand, is concerned with the coordination of material, information, and financial flows once the supply chain design is finalized. Effective supply chain strategies combine a range of approaches from operational flexibility (e.g., postponement, assemble-to-order (ATO), make-to-order (MTO), and leadtime (LT) reduction), channel alignment (e.g., contracts, vendor-managed inventories (VMI), and efficient consumer response initiatives (ECR)), and joint decision making through information deployment (e.g., point of sale (POS) data, collaborative planning forecasting and replenishment (CPFR), and schedule sharing).

Supply chain performance entails a trade-off between cost and service. While services is typically measured in terms of response time and fill rates, cost is captured in such metrics as average landed cost and total assets. The first step in managing this trade-off is to consider the nature of the demand faced by a company's products. Fisher (1997) classifies products into two broad categories: functional products and innovative products. *Functional products* include, for example, staples such as toothpaste that people buy in a wide range of retail outlets. Since such products satisfy basic needs, they have relatively stable predictable demand and long life cycles. This stability, in turn, invites competition, often resulting in low profit margins. To avoid low margins, many companies introduce innovations in technology or in fashion to lure customers (e.g., toothpaste with tartar control). Although profit margins of *innovative products* can be higher, the short life cycles and the great variety typical of these products make demand highly unpredictable.

Fisher (1997) asserts that with their high profit margins and volatile demand, innovative products require a fundamentally different supply chain design and management than stable low-margin functional products. The predictable demand of functional products makes market mediation easier. Companies can therefore afford to focus almost exclusively on minimizing physical cost, as the demand for functional products tends to be highly price sensitive. Hence, a *physically efficient* or *lean* supply chain is needed.

On the other hand, a *market-responsive* or *agile* supply chain is required for innovative products since the uncertain market demand increases the risk of shortages and excess supplies. While high profit margins and the importance of early sales in establishing market share for new product increase the cost of shortages, short life cycles increase the risk of obsolescence, hence the cost of excess supplies. Market mediation costs therefore dominate physical costs for such products.

In Fisher's view, the root cause of the problems plaguing many supply chains is a mismatch between the type of product and the type of supply chain. Such a mismatch can be avoided by devising an agile supply chain for innovative products while deploying a lean supply chain for functional products. Recently, a group of researchers (e.g., Katayama and Bennett, 1999; Naylor et al., 1999) have been trying to combine the advantages of lean and agile supply chains into a single concept termed "leagility".

Naylor et al. (1999) introduces the following definitions: "Agility means using market knowledge and a virtual corporation to exploit profitable opportunities in a volatile market place. Leanness means developing a value stream to eliminate all waste, including time, and to ensure a level schedule." Hence, leagility is defined as the capability of concurrently deploying the lean and agile paradigms. The use of market knowledge and the creation of an integrated supply chain are postulated as characteristics of equal importance under both paradigms. This is indeed consistent with our classification of approaches to supply chain management, namely, information deployment and channel alignment, respectively, as depicted in Fig. 1. LT compression, elimination of all waste, and rapid reconfiguration (quick changeover among product families) are also identified as key characteristics in the two paradigms, but with varying degrees of importance. We have classified such characteristics as operational flexibility. Finally, Naylor et al. asserts that an agile manufacturer must be robust in the sense that it must be able to withstand fluctuations and disturbances. Lean manufacturing, on the other hand, requires that demand variation be reduced so that the supply chain can be simplified and streamlined. This assertion is also consistent with Fisher's definition of market responsive versus physically efficient supply chains.

Naylor et al. further argues that the need for agility and leanness depends upon the total supply chain strategy, particularly by considering market knowledge and positioning of the decoupling point. The *decoupling point* separates the part of the supply chain that responds directly to the customer from the part of the supply chain that uses forward planning and a strategic stock to buffer against the variability in the demand of the supply chain. The positioning of the decoupling point therefore depends upon the longest LT, the end customer is willing to tolerate. This is also the point at which product variety increases significantly as variants of the end product fan out. Downstream from the decoupling point all products are pulled by the end customer; upstream from the decoupling point, the supply chain is typically forecast driven.

The authors conclude that the lean paradigm can be applied to the supply chain upstream of the decoupling point, as the demand is smooth and generic product flow through a number of value streams. The agile paradigm must be applied downstream from the decoupling point, as demand is variable and the product variety per value stream increases.

Research on leagility therefore hinges heavily on the identification of the decoupling point. This, in turn, is a manifestation of the postponement or delayed customization strategy for supply chain management. Postponement strategies, however, present a cross-functional challenge for implementation. The celebrated example of Hewlett-Packard's Deskjet printer (Lee et al., 1993) necessitated the redesign of the product (externalizing the power supply), the process (final assembly carried out at the distribution center) and the supply chain (introduction of patented packaging and palletizing process for ease of shipment and customization). Such three-dimensional concurrent engineering (Fine, 2000), in turn, could only be carried out through the participation of R&D, manufacturing, and logistics functions. In fact, any of the supply chain management approaches depicted in Fig. 1 necessitate the introduction of new organizational relationships, and require the support of key information and communication technologies. In other words, these initiatives cannot be readily deployed; they must be planned and rolled out following a carefully orchestrated implementation plan.

This paper discusses a tactical approach, transshipments, which can be implemented quickly with limited monetary investment. Furthermore, this solution can also provide a smooth transition toward the application of the above supply chain strategies. This tactical approach commonly practiced in multi-location inventory systems, involves monitored movement of stock between locations at the same echelon level of the supply chain. These stock movements are referred to as *transshipments*. Transshipments provide an effective mechanism for correcting discrepancies between the locations' observed demand and their available inventory. As a result, transshipments can lead to cost reductions and improved service without increasing inventory levels in the supply chain. Moreover, since transshipments enable the sharing of stock among locations, they naturally lead to coordinated replenishment policies across locations, thereby avoiding excessive procurement costs. In summary, transshipments can make a supply chain more agile while simultaneously enhancing its leanness.

The paper is organized as follows. Section 2 describes the role of transshipments not only as a readily deployable supply chain initiative, but also as an enabler for more strategic management initiatives described in Fig. 1. Section 3 introduces models for transshipments and briefly reviews the emerging literature on the topic. It further demonstrates through a series of illustrations how transshipments can simultaneously implement the lean and agile principles. Section 4 concludes the paper.

2. The role of transshipments

Traditionally, transshipments have been used as a tactical policy in multi-location inventory systems. The ability to transfer stock between locations at the same echelon level provides each location with an additional source of replenishment. Moreover, this additional source is typically characterized by short LT.

Using transshipments as an additional source of replenishment serves two goals. First, transshipments provide a secondary source of material when demand at a particular location turns out to be higher than expected while, in a neighboring location, excess stock is available. In such a case, the neighboring location can typically provide the required products faster than the original supplier. But once such emergency transfers are established and used regularly, the second goal, namely coordinating the replenishments, becomes attractive as well. That is, replenishment quantities are determined jointly, taking into consideration potential transshipments.

Coordinated replenishments between two (or more) locations require that a link be established between them. Such a link is part of the *supply chain design*. In particular, an appropriate communication channel, which enables information sharing, must be established in order to use the transshipment mechanism to its full extent. Since the establishment of communication channels between *any* two locations (in the same echelon level) may be too expensive, the supply chain designer needs to decide which communication channels are preferable to others. In particular, stocking locations can be divided into subgroups, where communication channels—hence, transshipments—are allowed *only within* these subgroups. An analysis of possible transshipment models can provide answers to such questions; such analysis is provided in Section 3.3.

From the point of view of *supply chain management*, transshipments support the range of innovative approaches mentioned in Section 1, for the coordination of material flows in the network. The relationship between transshipments and innovative supply chain strategies are discussed below.

One crucial problem in supply chain management is the long time delays that characterize replenishments. It is evident that LTs associated with transshipments are typically shorter; see Tagaras and Cohen (1992) and Tagaras (1999). With short LTs, the strategies of assemble to order and make to order become feasible, thereby enhancing the operational flexibility of the system.

According to the postponement strategy, the point of differentiation among products is delayed, so that they can be stored in a generic form until more information on the demand for the specific final products becomes known. The strategy of transshipments can be viewed in this context as physical postponement, where a specific product is associated with a specific location. Then, a product stored at one location can be transferred to another location, therefore satisfying the demand for the latter product type. As in postponement, the advantage here is the ability to transform a generic item (an item at any location) into a specific item (an item at a specific location) in a relatively short time. In Section 3.2, this point is further discussed, and illustrated.

Transshipments can be implemented with the help of immediate up-to-date information about the stock levels throughout the supply chain. This information is then available for implementing other strategies that require such information, e.g., VMI or efficient consumer response.

Transshipments eliminate time waste through LT reduction. This is especially appropriate in a volatile market place. Furthermore, as demand uncertainty grows, the benefits from transshipments are likely to be more significant (see, for example, Section 3.2, where the reduction in the replenishment quantity is shown to be proportional to the standard deviation of the demand). In addition, transshipments allow for the elimination of physical waste by keeping stock in a coordinated way among locations, thereby reducing the overall inventory in the supply chain. As such, transshipments contribute to satisfying the ideas of both the agile and lean concepts.

3. Leagility through transshipments

3.1. Modeling transshipments

In the literature on the transshipment problem, a differentiation is made between models that allow transshipments before demand is realized, typically referred to as *preventive* transshipments, and models that allow transshipments after demand is realized, typically referred to as *emergency* transshipments. The former group includes Allen (1958, 1962), Gross (1963), Karmarkar and Patel (1977), and Karmarkar (1979, 1981, 1987). The latter group includes Krishnan and Rao (1965), Tagaras (1989), Robinson (1990), and Herer and Rashit (1998, 1999). An interesting variation on these two extreme models is allowing transshipments while demand is being realized (Archibald et al., 1997). Tagaras and Cohen (1992) and Tagaras (1999) further consider non-zero replenishment LTs.

In his investigation of a two-location problem, Tagaras (1989) defines the complete pooling policy, whereby the amount transshipped from location i to location j is the minimum between the excess at location i and the shortage at location j. Under general cost structures, complete pooling is optimal. An order-up-to-S policy under complete pooling is also shown to improve customer service. Bertrand and Bookbinder (1998) extend the problem to a system whose stock keeping locations have non-identical costs. In particular, they consider a warehouse following a periodic order-up-to-S policy based on the system stock. Once the warehouse receives a shipment, it is entirely allocated to the retailers, who experience independent and (not necessarily identically distributed) normal demand. Prior to a new replenishment (order by the warehouse), system stock is redistributed—in a preventive transshipment mode—among the retailers to minimize the expected holding, backorder and transshipment costs. In the case with identical retailers, Bertrand and Bookbinder analytically show that redistribution reduces the variance of the net inventory prior to a new order. For the case with non-identical retailers, a one-parameter-at-a-time simulation experiment shows that higher values of the length of the replenishment cycle, the number of retailers, holding costs, LTs from the warehouse to the retailers, coupled with low values for transshipment costs, supplier LTs, and shortage penalties, favor a redistribution policy.

Evers (1997) considers emergency transshipments and conclude that the resulting benefits lie not only in the pooling of demands, but also in the pooling of LTs. Evers (1999) identifies the links between transshipments and order splitting, and measures the effects of pooling on relevant cost drivers. Needham and Evers (1998) investigate under which combination of cost parameters pooling or partial pooling (in the form of transshipments) should be employed. Their model, which incorporates replenishment and transshipment LTs, also includes emergency resupply from the higher echelon. They conclude that the stock-out cost is the major factor determining whether transshipments should be employed. Alfredsson and Verrijdt (1999) consider a two-echelon inventory system for service parts. They assume a one-for-one replenishment strategy with emergency lateral transshipments as well as direct delivery from the warehouse and from the plant. They present an approximate model that is used to calculate relevant performance measures, and verify the accuracy of this model through simulation. The results obtained through simulation demonstrate that the performance of the inventory system is insensitive to the replenishment LT distribution and that lateral transshipments and direct deliveries can lead to significant cost savings. They conclude that supply flexibility always pays off.

Tagaras and Vlachos (2002) consider a two-retailer configuration with identical cost structures under non-negligible transshipment LTs and different demand processes. The key motivation for this analysis is that the items in transit cannot be used to satisfy demand at any of the retailers; hence, transshipments may actually lead to system wide service deterioration. Through simulation, they study the impact of preventive lateral transshipments through a partial pooling policy, where a location is willing to transship inventory that is above and beyond the local target inventory level. They conclude that transshipments are significantly beneficial only with highly variable demand. Alternative stock control policies do not change the above conclusion. Herer and Rashit (1999) examine the effects of fixed replenishment costs in the single-period problem. Rudi et al. (2001) examine the differences between local decision making and central decision making in a two-echelon supply chain, and demonstrate the superiority of coordinated decision making.

Herer and Rashit (1998) show that Tagaras's complete pooling policy does not guarantee optimality under all cost structures. Herer and Tzur (2001) are the first researchers to investigate the strategy of transshipments in a dynamic deterministic demand environment over a finite planning horizon. In a model that includes both fixed and variable transshipment costs, they develop a variant of the Wagner–Whitin algorithm to find an optimal policy.

In models where transshipments occur after demand is realized, the period starts when every retailer has a certain amount of stock. The first event in the period is the occurrence of demand. After observing the demand, transshipments are made and then demand is satisfied. At that point, backlogs and inventory are observed, penalty and holding costs, respectively, are incurred. Finally, using a base stock policy, replenishment orders are placed, received immediately, and are used to satisfy the backlog. The remaining inventory is carried to the beginning of the next period.

The decision variables in this problem are the transshipment and replenishment amounts. Hence, we use X_{ij} to denote the transshipment quantity from retailer i to retailer j. We further denote by S_i the base stock level at retailer i. We also use the auxiliary variable I_i to represent the inventory level at retailer i after transshipments and demand fulfillment. Then, given the base stock level, S_i, and the observed demand, d_i, the inventory level I_i can be expressed as

$$I_i = S_i - d_i - \sum_{j=1}^{N} X_{ij} + \sum_{j=1}^{N} X_{ji}.$$

Letting $I_i^+ = \max\{I_i, 0\}$ and $I_i^- = \max\{-I_i, 0\}$, the total cost of the system in a given period is given by

$$\text{Cost} = \sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij}X_{ij} + \sum_{i=1}^{N} h_i I_i^+ + \sum_{i=1}^{N} p_i I_i^-.$$

3.2. Two identical locations

We first show that transshipments can be used to improve service, one of the cornerstones of agility, and/or reduce stocks, one of the cornerstones of leanness. We begin by considering two identical stock-keeping locations. The locations are identical in the sense that demand at each location is assumed to be distributed normally with mean μ and standard deviation σ. Table 1 summarizes the performance of this system under different materials management strategies. Let us start our analysis by considering each location independently, i.e., without any transshipments

Table 1
System performance under various supply chain strategies

Strategy	Individual location's order quantity	Individual location's no-stockout probability	System wide no-stockout probability
No transshipments	$\mu + 1.282\sigma$	0.900	0.810
Transshipments—agile	$\boldsymbol{\mu + 1.282\sigma}$	0.978	0.965
Transshipments—lean (system-based)	$\mu + 0.621\sigma$	0.869	**0.810**
Transshipments—lean (location-based)	$\mu + 0.741\sigma$	**0.900**	0.853
Transshipments—leagile	$\mu + 0.906\sigma$	0.933	0.900

between the two locations. This means that we have two independent newsboy problems, where each location has the same critical fractile. For the purpose of illustration, assume that this critical fractile is 0.9; that is, the no-stockout probability at each location is 90%. This value was chosen solely for illustrative purposes; all the results remain qualitatively unchanged as long as this critical fractile is >0.5. Since the demand is normally distributed the optimal order-up-to point for each location can be identified by a critical factor, $k_{0.9} = 1.282$ (independent of the location), such that the optimal order-up-to quantity for location i is $\mu + 1.282\sigma$.

Without transshipments each location has a no-stockout probability of 0.9, but the system, which consists of the two stocking locations, has a system wide no-stockout probability of $0.9^2 = 0.81$. Note that, while the locations are considered together, the inventory is still split between the two stocking locations. Transshipments can be used to improve the overall system's performance as well as the performance of each location. Consider complete pooling between the two locations without changing the order-up-to point; that is, the amount transshipped is the minimum between the excess inventory at the sending location and the shortage at the receiving location. In this case, the combined demand would be normally distributed with mean 2μ and standard deviation of $\sqrt{2}\sigma$. The combined order quantity is $2(\mu+1.282\sigma)$ resulting in a system wide no-stockout probability of 0.965. Each individual location's no-stockout probability is actually higher than this because it is possible that one location has a surplus while the other location has a shortage, but the surplus is not enough to satisfy the shortage. In this case, the first location would not have a stockout while the second location would (and of course the system as a whole would have a stockout). Each location's no-stockout probability is given by Proposition 1 in Tagaras (1989) and in this case is equal to 0.978.

Thus far, we have illustrated how transshipments can be used to improve service, hence, enhance agility. Conversely, we can consider transshipments as a means of reducing inventories to enhance leanness. If we want the new system's no-stockout probability to be the same as the original system (0.81), then each location would need to order $\frac{1}{2}(2\mu + k_{0.81}(\sqrt{2}\sigma)) = \mu + \frac{1}{2}\sqrt{2}(0.878)\sigma = \mu + 0.621\sigma$ (instead of $\mu + 1.282\sigma$). This gives each individual location a no-stockout probability of 0.869. If instead of equating the old and new system based on the system wide no-stockout probabilities, we equate them with respect to each individual location's no-stockout probability then each location would have to order $\mu + 0.741\sigma$ to insure a no-stockout probability of 0.9 at each location after transshipments. This yields a system wide no-stockout probability after transshipments of 0.853.

Finally, if we try to enhance both leanness and agility simultaneously, we can both reduce inventory and improve service (lean and agile), but of course to a smaller extent than we can affect each individually. We may choose to equate the no-stockout probabilities at each individual location without transshipments with the system wide no-stockout probability after transshipments. This means that each location would have to order $\frac{1}{2}(2\mu + k_{0.90}(\sqrt{2}\sigma)) = \mu + 0.906\sigma$ and each individual's no-stockout probability would be 0.933.

3.3. Multiple locations

In Section 3.2, we illustrated how transshipments can help achieve leagility in the framework of a two-location problem. When the number of locations grows, the cost of having *all* locations possibly transship to each other is often prohibitive. In this section, we therefore examine the problem of arranging locations into groups, where only intra-group transshipments are allowed. Many of the benefits of transshipments can be obtained by pooling (in a transshipment sense) even a small number of locations, while reducing the complexity of managing the associated material movement. This is demonstrated in Figs. 2 and 3. These figures represent a 120-location problem. Again we examine the case where each location's demand is normally distributed with mean μ and standard deviation σ, and the stocking level at each location is initially set so that the pre-transshipment no-stockout probability is 90%, i.e., an order-up-to quantity of $\mu + 1.282\sigma$. In both figures, the x-axis represents the number of locations put together in a pooling group. Thus a value of one represents no transshipments, while a value of ten represents twelve pooling groups, each containing ten locations.

In Fig. 2, we hold the stocking level at each location constant and observe how the system wide no-stockout probability changes. We see that even a grouping of four location increases the system wide no-stockout probability from nearly zero (without transshipments) to over 80%. In Fig. 3, we hold the system wide no-stockout probability constant (and equal to 0.9^{120}) and observe how the system wide order-up-to quantities

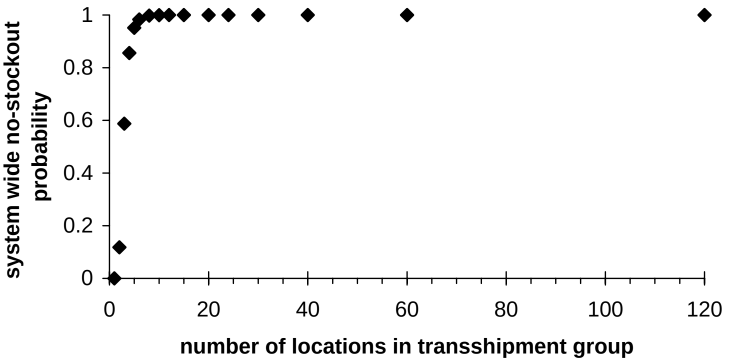

Fig. 2. Effect of size of transshipment groups, the agile outlook.

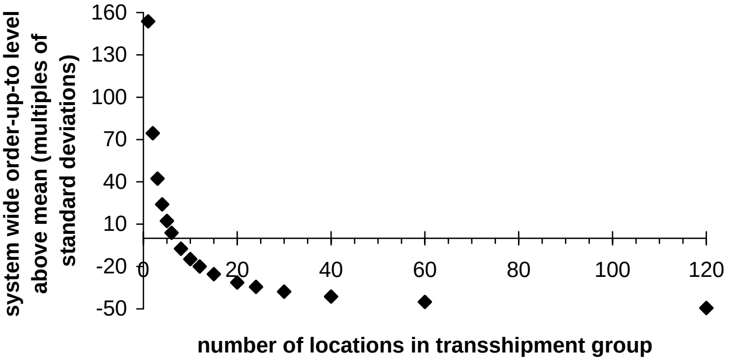

Fig. 3. Effect of size of transshipment groups, the lean outlook.

change. Again the figure shows that a grouping of four locations drastically reduces the system wide order-up-to levels.

Consider now a nine-location problem, where the challenge is to split the locations into three groups of three. Tagaras (1999) also considered this problem, albeit with replenishment LTs. After we examine this problem, we will qualitatively compare our results to his. Our only requirement is that, after pooling, the service level (expressed as the no-stockout probability) of each transshipment group is equal to α. As before, α is assumed to be >0.5. The demand at each location $i, i = 1, \ldots, 9$, is distributed normally with mean μ_i and standard deviation σ_i. As in Tagaras (1999), we examine the case where the first six locations (locations 1–6) are identical and the last three locations (locations 7–9) are identical, though similar results are available without this assumption. Since demand is normally distributed, the optimal order-up-to point for any group of locations, g, with mean μ^g and standard deviation σ^g can be identified by a critical factor, k_α, such that the optimal order-up-to quantity is $\mu^g + k_\alpha \sigma^g$.

We would like to compare the performance of groups with non-identical locations (Grouping 1) (e.g., $G_{11} = \{1,2,7\}$, $G_{12} = \{3,4,8\}$, and $G_{13} = \{5,6,9\}$) to the performance of groups consisting of identical locations (Grouping 2) (e.g., $G_{21} = \{1,2,3\}$, $G_{22} = \{4,5,6\}$, and $G_{23} = \{7,8,9\}$). In other words, we examine whether we should collect similar retailers together, thus creating dissimilar groups or create similar groups, each containing dissimilar retailers. Since we are interested in the no-stockout probability of the groups, only the order quantities of the groups, and not of the individual retailers, are of interest. We will let the order-up-to quantity of group G_{ij} be represented by S_{ij}. The order-up-to points are thus

$$S_{11} = S_{12} = S_{13} = 2\mu_1 + \mu_7 + k_\alpha\sqrt{2\sigma_1^2 + \sigma_7^2},$$

$$S_{21} = S_{22} = 3\mu_1 + k_\alpha\sqrt{3}\sigma_1$$

and

$$S_{23} = 3\mu_7 + k_\alpha\sqrt{3}\sigma_7.$$

We would like to determine which of these two groupings uses less inventory to achieve the required service level. This relationship is investigated in the next proposition.

Proposition 1. *Grouping 1 uses more inventory than Grouping 2 to obtain the same service level, namely,* $S_{11} + S_{12} + S_{13} \geqslant S_{21} + S_{22} + S_{23}$.

Proof. We prove the proposition through a series of inequalities. We need to show

$$S_{11} + S_{12} + S_{13} \geqslant S_{21} + S_{22} + S_{23}. \tag{1}$$

Substituting the actual values, we obtain

$$6\mu_1 + 3\mu_7 + 3k_\alpha\sqrt{2\sigma_1^2 + \sigma_7^2} \geqslant 6\mu_1 + 3\mu_7 + 2k_\alpha\sqrt{3\sigma_1^2} + k_\alpha\sqrt{3\sigma_7^2}. \tag{2}$$

Simple algebra yields (note that k_α is positive since we assumed that the desired no-stockout probability is at least 0.5)

$$3\sqrt{2\sigma_1^2 + \sigma_7^2} \geqslant 2\sqrt{3\sigma_1^2} + \sqrt{3\sigma_7^2}. \tag{3}$$

Squaring both sides of inequality (3), we obtain

$$9(2\sigma_1^2 + \sigma_7^2) \geqslant 12\sigma_1^2 + 12\sqrt{\sigma_1^2\sigma_7^2} + 3\sigma_7^2. \tag{4}$$

Simplifying,

$$\sigma_1^2 - 2\sqrt{\sigma_1^2\sigma_7^2} + \sigma_7^2 \geqslant 0, \tag{5}$$

$$\left(\sqrt{\sigma_1^2} - \sqrt{\sigma_7^2}\right)^2 \geqslant 0. \tag{6}$$

Clearly inequality (6) holds, which completes the proof of the proposition. □

Corollary 2. *The result of Proposition 1 holds for any value of the mean demand* (μ_i) *for each of the nine locations.*

The corollary follows from the fact that the proof of Proposition 1 holds without modification. Intuitively, this corollary is very important. Whereas Proposition 1 demonstrates that "similar" locations should be put together in transshipment groups, Corollary 2 explains what is meant by "similar" locations. That is, locations with similar standard deviations of demand should be put together. At first, this result may appear counterintuitive. One might claim that more variable locations should be pooled with more stable locations so that the latter's stability could compensate for the former's variability. The problem with this argument is that the latter's stability causes the location to order-up-to almost exactly to its mean demand rendering it unable to help with the former location's variability. When similar locations are put together, the stable locations do not need to order much extra stock and the well-known law of large numbers "reduces" the joint variability of the variable locations, thus resulting in lower overall inventories.

Here we have investigated the question of how to split nine locations into three groups of three. As we have mentioned earlier, Tagaras (1999) considered the same question in a different setting. His model includes a delivery (replenishment) delay from the central supplier to the individual locations. He found that, in one instance, the cost of Grouping 1 is 15% higher than the cost of Grouping 2. Similarly he found that the "safety stock" of Grouping 1 is 18% higher. These findings are analogous to our Proposition 1. Tagaras (1999) and Tagaras and Vlachos (2002) also note results analogous to our Corollary 2.

3.4. Leagility findings in the literature

Even though leagility is a new term (introduced in 1999), the literature on transshipments goes as far back as the 1950s, predating both the lean and agile concepts. This literature is rich on examples which show that transshipments achieve leagility. In this section, we will examine examples from Krishnan and Rao (1965), Tagaras (1989), Tagaras (1999), and Herer and Rashit (1999).

Example 1 in Krishnan and Rao (1965) shows that, in a seven-location setting where each location has identical cost parameters, the cost-minimizing levels of inventory decrease at each of the locations whenever another location is added.

Tagaras (1989) presents one example for which three policies are compared: the optimal no transshipment policy, the cost-minimizing policy with transshipments, and the cost-minimizing policy subject to the fill rates being at least 99% at each location. His results are summarized in Table 2.

As can be seen in Table 2 when going from the no-transshipment policy to the cost-minimizing policy, both the total order-up-to quantities are reduced and the performance measures (fill rate and no-stockout probability) are improved simultaneously. Thus, transshipments, in this example, yield a lower-cost, leaner, and more agile system. If we wish to emphasize agility by constraining the system to have very high service levels (a fill rate of 99% at each location), then we have to compromise. The total cost ends up being somewhere between the total cost of the other two policies, but the total order-up-to-quantity is higher than it is without transshipments.

Tagaras (1999) examines the effects of transshipments on the cost-minimizing policy in the presence of replenishment LTs from the original supplier while transshipments are still assumed to have negligible LTs. In each of the 60 examples examined by Tagaras (1999) (Table 5 therein), introducing transshipments reduced the total safety stock, reduced the total cost, and increased the fill rate. Thus, we see that transshipments can also help to obtain the goal of leagility even in the presence of LTs.

Another way transshipments help with agility is by drastically reducing the time it takes to respond

Table 2
Summary of the example from Tagaras (1989)

Strategy	Total order-up-to quantity	Total cost	Average fill rate	Average no-stockout probability
No transshipments	424.5	128.31	0.971	0.792
Transshipments—cost minimization	411.7	110.07	0.982	0.872
Transshipments—minimum fill rate	432.8	113.85	0.990	0.925

to a shortage in the system. Without the replenishment LTs, the effects of a particularly large demand are felt only until the following period. With LTs this effect can last as long as the LT itself. When the LTs of the locations are equal, transshipments can only help by reducing the shortages associated with this large demand. However, when the LTs of the locations are unequal and the large demand is manifested at a location with a long LT, transshipments can be used to reduce the time it takes to eliminate the effects of the large demand from the system. Even though this was not directly investigated by Tagaras (1999), where only order-up-to policies are considered, the phenomenon is nonetheless clear. Examining the safety factors associated with the order-up-to point of each location (Table 3 therein), we observe that the safety factors are almost identical whether all locations have a LT of one period or a LT of three periods (Table 1 therein present the examples in detail). However, when the LT of one of the locations is three periods and the LT of the other two locations is one period, then the safety factor at the location with the longer LT actually *goes down*, while the safety factor at the other locations goes up. In our view, this is because the locations with the shorter LTs can more quickly respond to a large demand at the long LT location than the location itself. Thus, it is actually worthwhile to have the safety for the long LT location located at the short LT locations.

Finally, Herer and Rashit (1999) examine the single-period problem with transshipments in the presence of fixed replenishment costs. Let us consider their Example 2 (Table 1 therein). For this problem, when initial inventories are zero, the cost-minimizing policy without transshipments is to order 140.25 units (Table 3 therein). Contrast this quantity with 132.31 units when transshipments are allowed. An interesting aspect of the adaptability of the system with transshipments is apparent when the initial inventories in some period are non-zero (case (d) of Table 3 therein). In this situation, the best policy without transshipments is for location 2 to order (location 1 does not order), while the best policy taking into account transshipments is for location 1 to order (location 2 does not order). This adaptability reduces the cost of the cost-minimizing policy by 35%. The majority of this benefit is not due to transshipments alone (which reduce the cost in this example by only 4%), but rather it is due to the coordination made possible by knowing that transshipments are available. This allows the system to put the inventory where it is needed.

4. Summary

Leagility is a recently introduced concept aimed at creating both lean and agile supply chains. Research on leagility relies on postponement (or delayed customization) of the product so as to identify an appropriate decoupling point along the supply chain. Postponement, however, typically necessitates modifications in product and process designs as well as in organizational relationships. Such modifications may be costly and time consuming. In this paper, we discussed transshipments as a fast and inexpensive approach to concurrently reduce cost and improve service in a supply chain. In particular, we have demonstrated, through a series of examples, how transshipments can decrease cost by reducing the overall inventory levels and improve service by both reducing the no-stockout probability and shortening the replenishment LTs. In the case of non-identical retailers, we have also demonstrated how the effectiveness of the transshipment strategy can be further improved by dividing the retailers into particular subgroups.

Acknowledgements

This research was partially funded by a grant from AFIRST, Association Franco-Israélienne pour la Recherche Scientifique et Technologique.

References

Alfredsson, P., Verrijdt, J., 1999. Modeling emergency supply flexibility in a two-echelon inventory system. Management Science 45, 1416–1431.

Allen, S.G., 1958. Redistribution of total stock over several user locations. Naval Research Logistics Quarterly 5, 549–571.

Allen, S.G., 1962. A model with setup charge. Management Science 8, 99–108.

Archibald, T.W., Sassen, A.A.E., Thomas, L.C., 1997. An optimal policy for a two-depot inventory problem with stock transfer. Management Science 43, 173–183.

Bertrand, L.P., Bookbinder, J.H., 1998. Stock redistribution in two-echelon logistics systems. Journal of the Operational Research Society 49, 966–975.

Camm, J.D., Chorman, T.E., Dill, F.A., Evans, J.R., Sweeney, D.J., Wegryn, G.W., 1997. Blending OR/MS, judgment, and GIS: Restructuring P&G's supply chain. Interfaces 27, 128–142.

Evers, P.T., 1997. Hidden benefits of emergency transshipments. Journal of Business Logistics 18, 55–76.

Evers, P.T., 1999. Filling customer orders from multiple locations: a comparison of pooling methods. Journal of Business Logistics 20, 121–139.

Fine, C.H., 2000. Clockspeed-based strategies for supply chain design. Production and Operations Management 9.3, 213–221.

Fisher, M., 1997. What is the right supply chain for your product? Harvard Business Review (March–April), 105–116.

Gross, D., 1963. Centralized inventory control in multi-location supply systems. In: Scarf, Gilford, Shelly (Eds.), Multi-Stage Inventory Models and Techniques, Stanford University Press, Palo Alto, CA.

Herer, Y.T., Rashit, A., 1998. Policies in a general two-location infinite horizon inventory system with lateral stock transshipments. Working Paper, Department of Industrial Engineering, Tel Aviv University, Israel.

Herer, Y.T., Rashit, A., 1999. Lateral stock transshipments in a two-location inventory system with fixed and joint replenishment costs. Naval Research Logistics 46, 525–547.

Herer, Y.T., Tzur, M., 2001. The dynamic transshipment problem. Naval Research Logistics 48, 386–408.

Karmarkar, U.S., 1979. Convex stochastic programming and multilocation inventory problem. Naval Research Logistics Quarterly 26, 1–19.

Karmarkar, U.S., 1981. The multiperiod multilocation inventory problem. Operations Research 29, 215–228.

Karmarkar, U.S, 1987. The multilocation multiperiod inventory problem: bounds and approximations. Management Science 33, 86–94.

Karmarkar, U.S., Patel, N., 1977. The one-period *N*-location distribution problem. Naval Research Logistics Quarterly 24, 559–575.

Katayama, H., Bennett, D., 1999. Agility, adaptability and leanness: A comparison of concepts and a study of practice. International Journal of Production Economics 60–61, 43–51.

Krishnan, K.S., Rao, V.R.K., 1965. Inventory control in *N* Warehouses. Journal of Industrial Engineering 16, 212–215.

Lee, H.L., Billington, C., Carter, B., 1993. HP gains control of inventory and service through design for localization. Interfaces 23, 1–11.

Naylor, J.B., Naim, M.N., Berry, D., 1999. Leagility: Integrating the lean and agile manufacturing paradigms in the total supply chain. International Journal of Production Economics 63, 107–118.

Needham, P.M., Evers, P.T., 1998. The influence of individual cost factors on the use of emergency transshipments. Transportation Research 34, 149–160.

Robinson, L.W., 1990. Optimal and approximate policies in multiperiod multilocation inventory models with transshipments. Operations Research 38, 278–295.

Rudi, N., Kapur, S., Pyke, D., 2001. A two-location inventory model with transshipment and local decision making. Management Science 47, 1668–1680.

Tagaras, G., 1989. Effects of pooling in two-location inventory systems. IIE Transactions 21, 250–257.

Tagaras, G., 1999. Pooling in multi-location periodic inventory distribution systems. Omega 27, 39–59.

Tagaras, G., Cohen, M., 1992. Pooling in two-location inventory systems with non-negligible replenishment lead times. Management Science 38, 1067–1083.

Tagaras, G., Vlachos, D., 2002. Effectiveness of stock transshipment under various demand distributions and non-negligible transshipment times. Production and Operations Management 11 (2).